

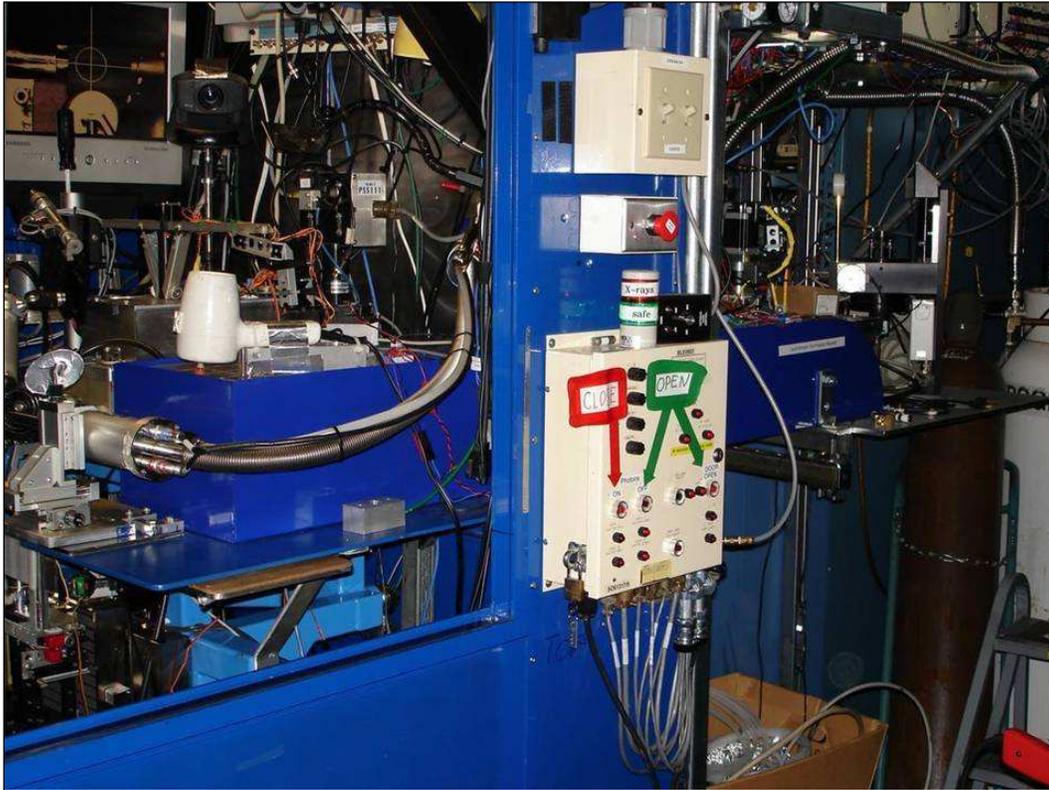
# **Increasing the structure determination success rate**

## **1 Screening infrastructure**

- a) offline targeting and unattended execution
- b) Room temperature tray screening

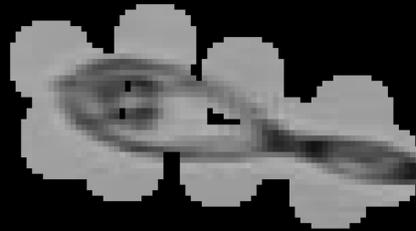
## **2 Crystal triage technology**

- a) Simulate the experiment to discover minimum required data quality
- b) Geometric strategy for unpredictable damage



Cryogenic sample conveyor labyrinth installed in beamline 8.3.1. Samples can be inserted into the external port of the conveyor (right) and moved into the robot-accessible internal port (left) without ever turning off the x-rays and disrupting ongoing data collection.

# Pepper Screening

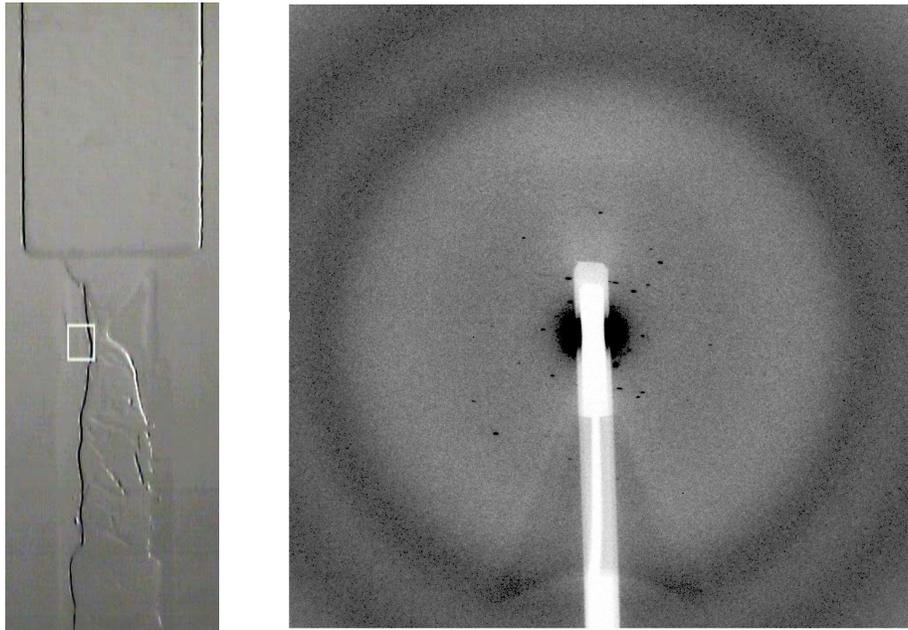


Samples that have not been pre-targeted may be subjected to “pepper screening”. Here the silhouette of the loop is taken as a collection of potential targets. First, the centroid of the silhouette is selected as a target and the pixels corresponding to the exposed area are eliminated as potential targets. Next, the new centroid is taken as the next target. This process is repeated until no pixels are left. The result is a self-avoiding series of targets that achieves maximum coverage of the loop in the least amount of shots.

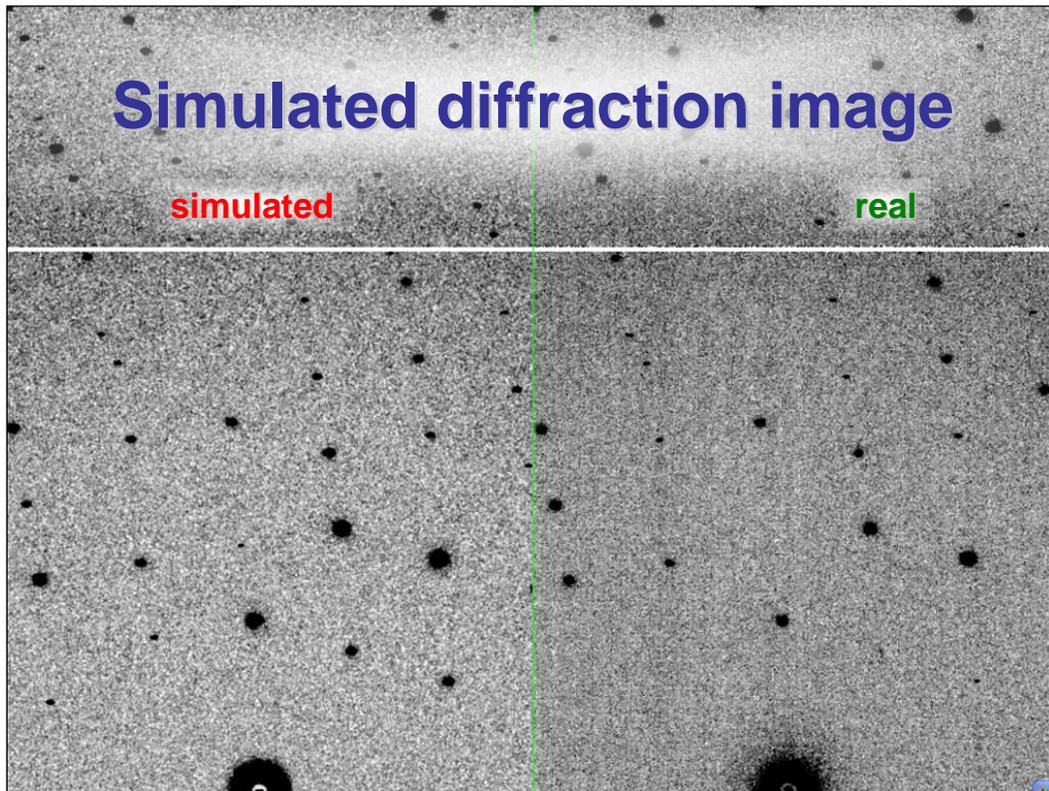


The non-rotating “plate goniometer” (left) was developed last year. A new model for taking rotation-camera data is now being developed (right). Either of these devices will accept most any crystallization tray that can be rotated 90 degrees and probe arbitrary points in each crystallization drop with a 50  $\mu\text{m}$  x-ray beam. Installing a plate goniometer takes  $\sim 30$  minutes and the location  $\sim 300$  mm downbeam from the “normal” protein crystallography sample mounting point makes this technology transferable to most any beamline.

*in-situ* screening: A11-1 target 1



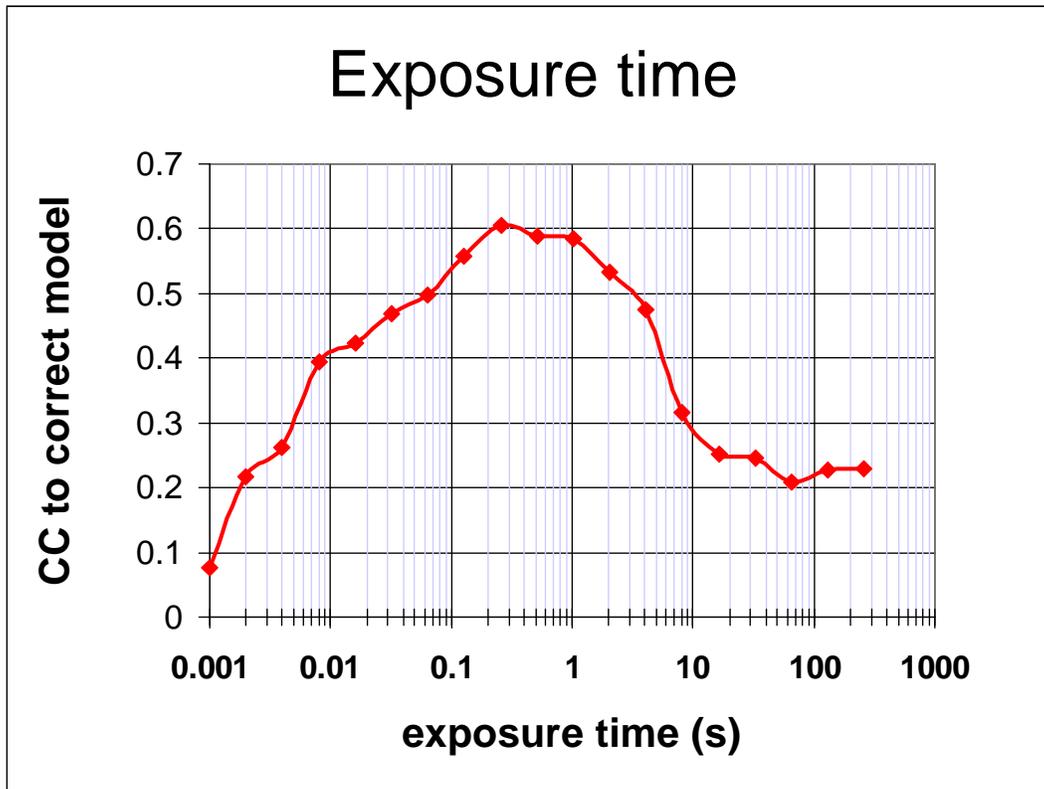
Demonstrates that diffracting crystal “hits” in a screening setup may not resemble crystals at all when examined optically (left), but still produce a clearly crystalline diffraction pattern (right).



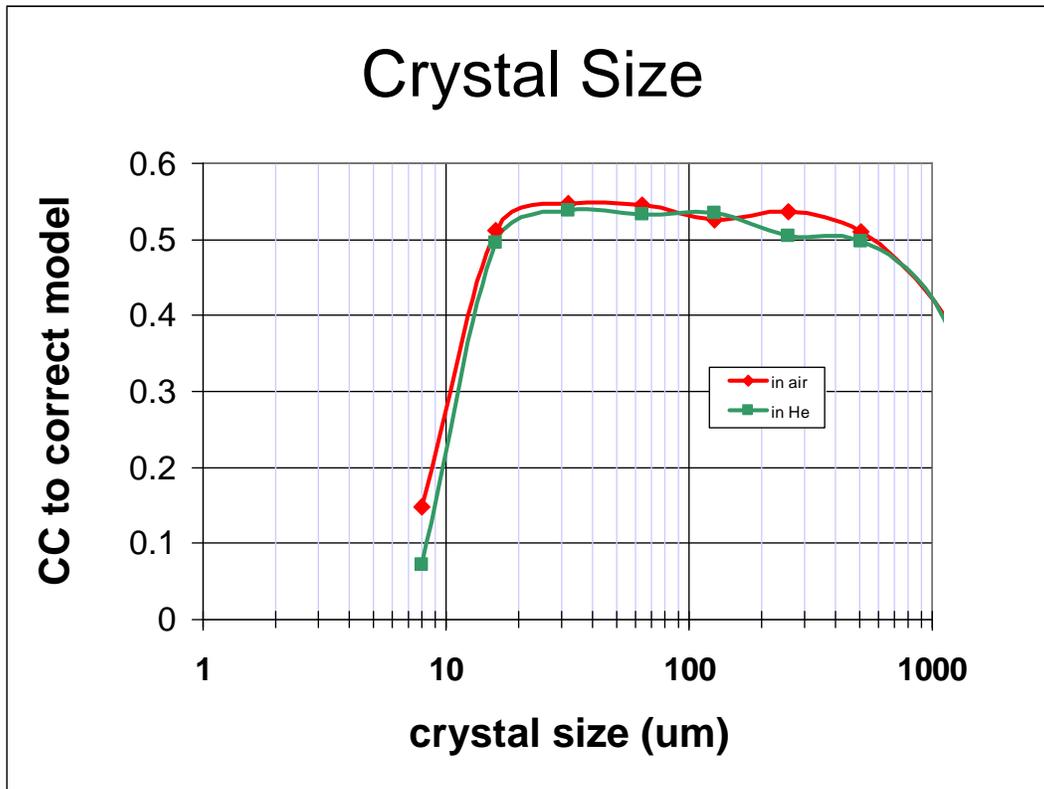
Side-by-side comparison of a diffraction image (right) next to a simulation of the same experiment (left). This simulation is based entirely on first principles. The characteristics of the detector, crystal, and x-ray beam are all modeled on an absolute scale.

Simulated	statistic	Real
5.4%	$R_{\text{merge}}$	6.2%
18.9	I/sd	16.2
1.9	I/sd (1.4 Å)	1.6
1.7 4.0 0.020	SDCORR	1.0 2.2 0.065
36.8	PADFPH	31.46
3.871	mlphare f''	3.476
0.178	FOM	0.192
0.545	FOMDM	0.664
0.6270	CC(1H87)	0.6090
0.170	$R_{\text{cryst}}$	0.184
0.205	$R_{\text{free}}$	0.231

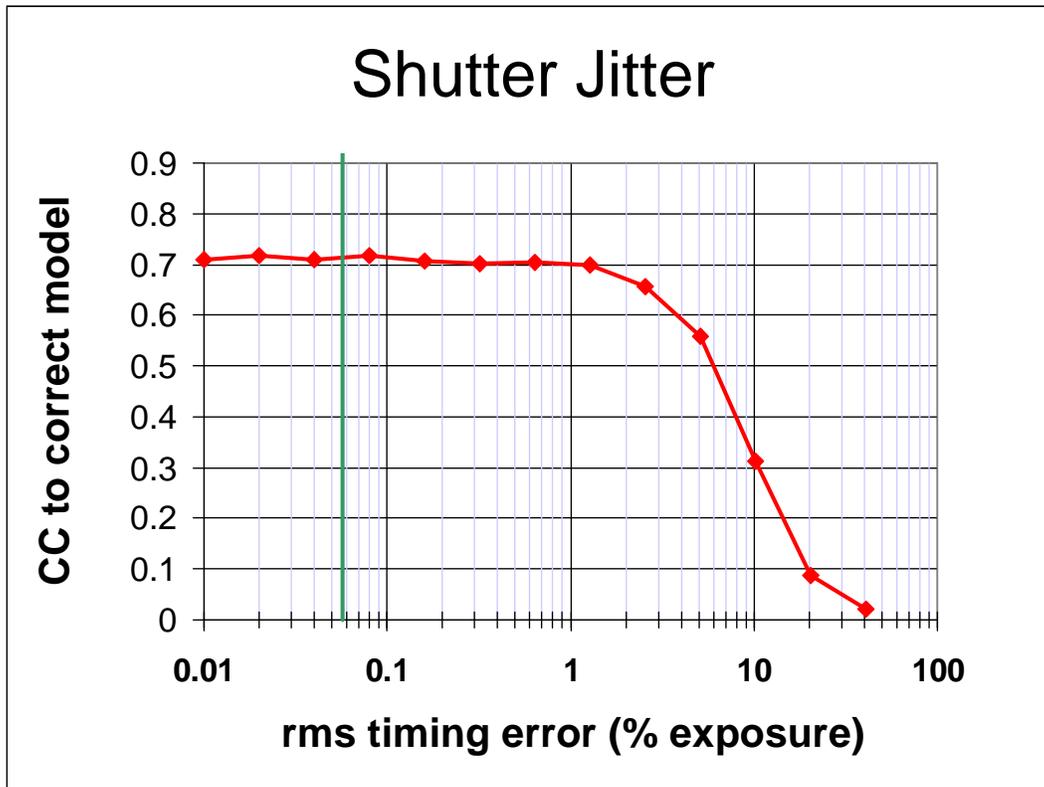
Results of processing the images from the simulation next to the results of processing data collected from the real-world experiment that was modeled in the simulation. The simulation produces slightly better data than reality, but most of the statistics are remarkably close. SDCORR is the optimised error model derived in the program SCALA. PADFPH is the maximum phased anomalous difference Fourier peak height. FOM is the figure of merit from MLPHARE. FOMDM is the figure of merit from the program DM. CC(1H87) is the correlation coefficient of the map from DM to the published model (PDB id 1H87).



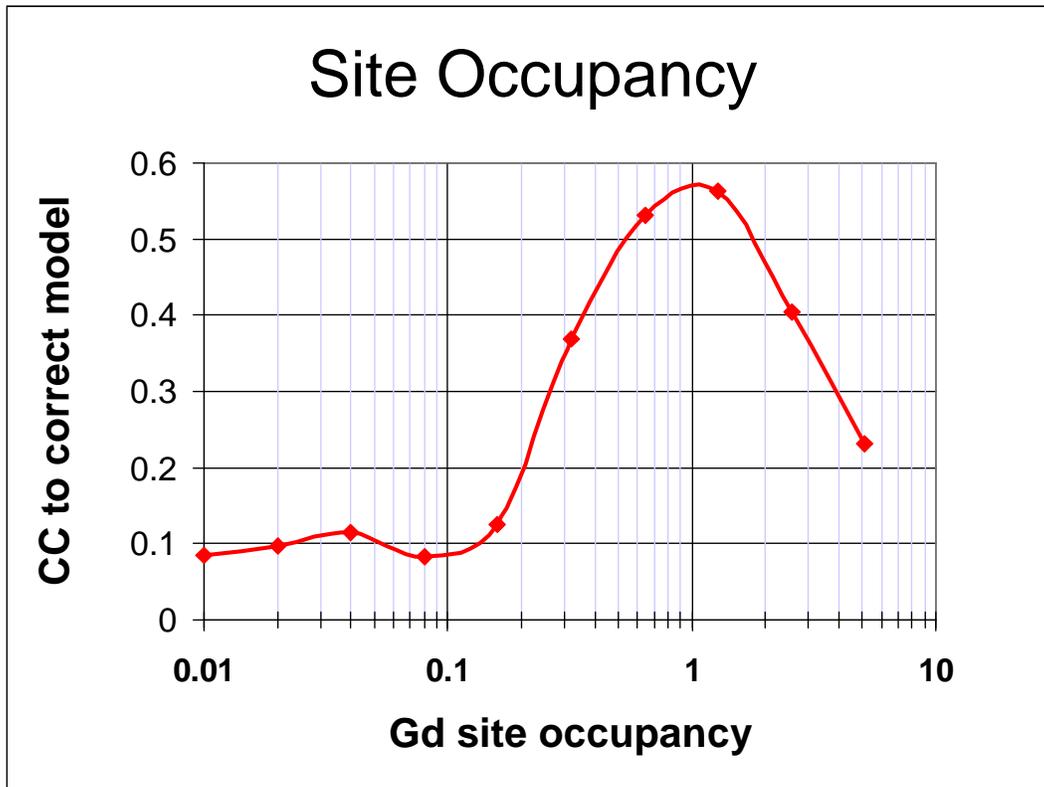
Results from adjusting the exposure time in the simulation. The exposure time in the real data set was 0.1s. Note that there is an optimal exposure time and very long exposures are detrimental. This effect is due to overloads.



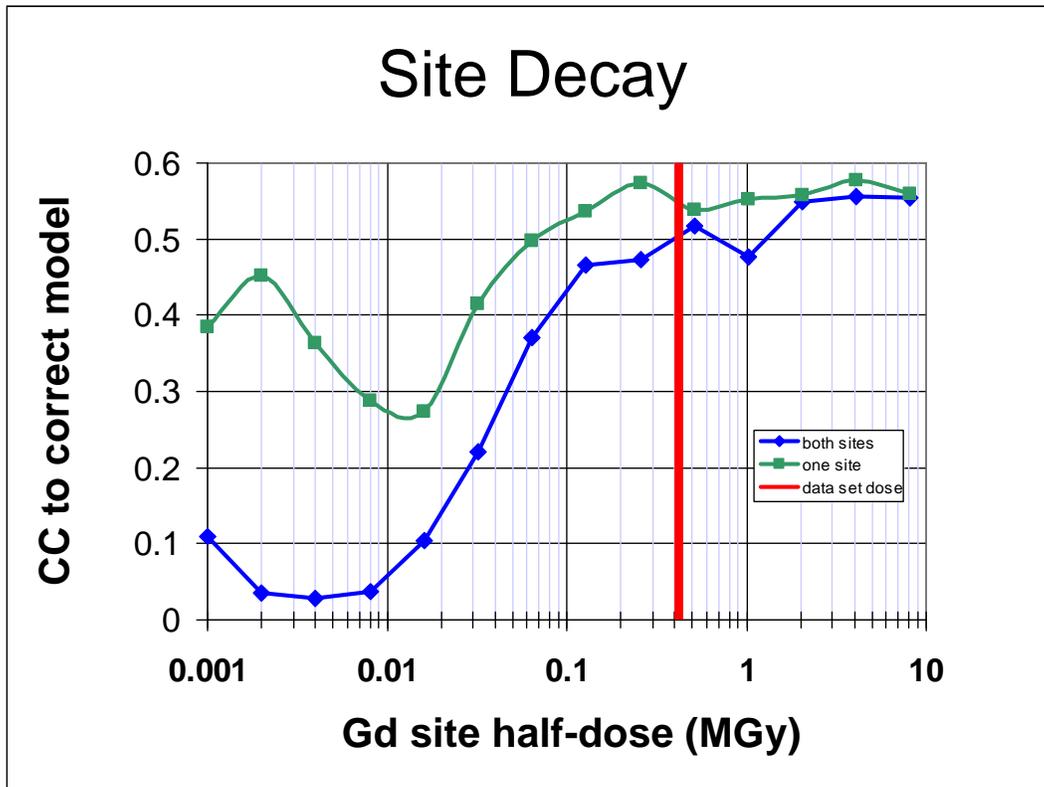
Results from changing the size of the crystal in the simulation. for each of these points the beam size was adjusted to match that of the crystal and the exposure time was also adjusted to put all the images on the same scale. However, since a realistic radiation damage rate is modeled, the very small crystals decay too rapidly for the simulated data set to be solved. Note that using a He path is of limited usefulness.



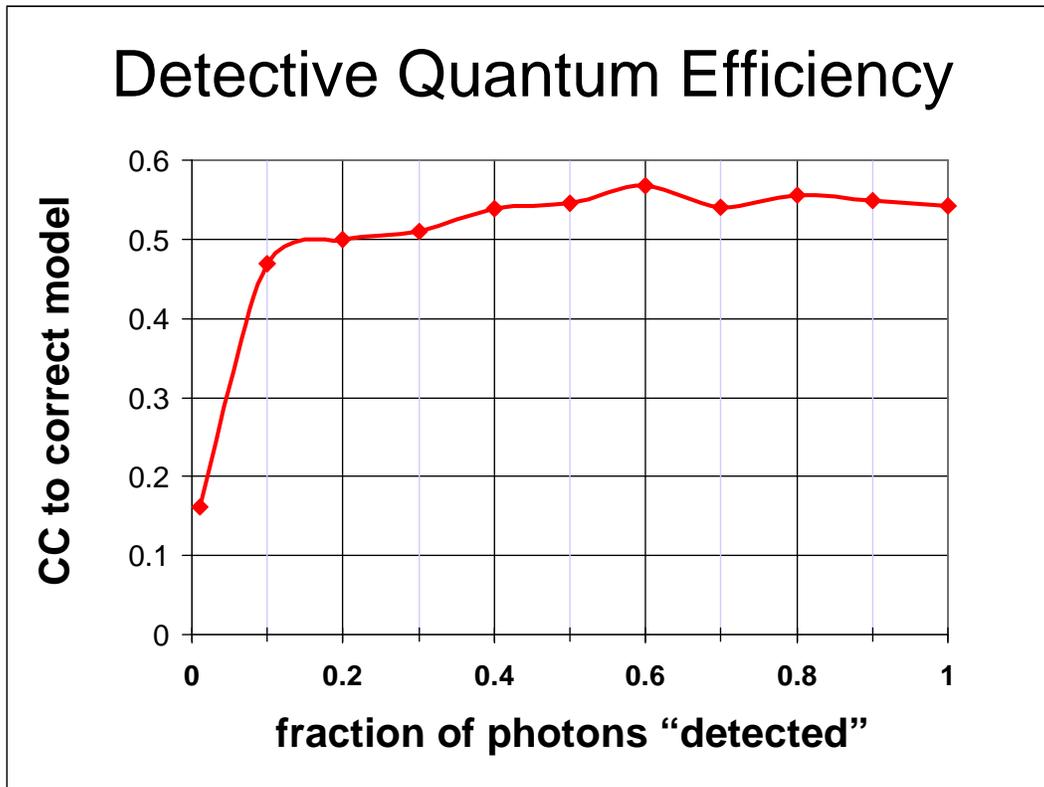
Sources of noise in the beamline can be detrimental to MAD/SAD. In this case, the Bijvoet ratio  $\langle \Delta F \rangle / \langle F \rangle$  was 3.5%. Notice that the “crossover” point of this curve occurs when the noise from the shutter amounts to 3.5%. Note also that shutter noise of less than 1% error has the same final structure quality as no shutter noise at all. This sigmoidal behavior is typical of these simulations and illustrates how small signals are expected to interact with noise.



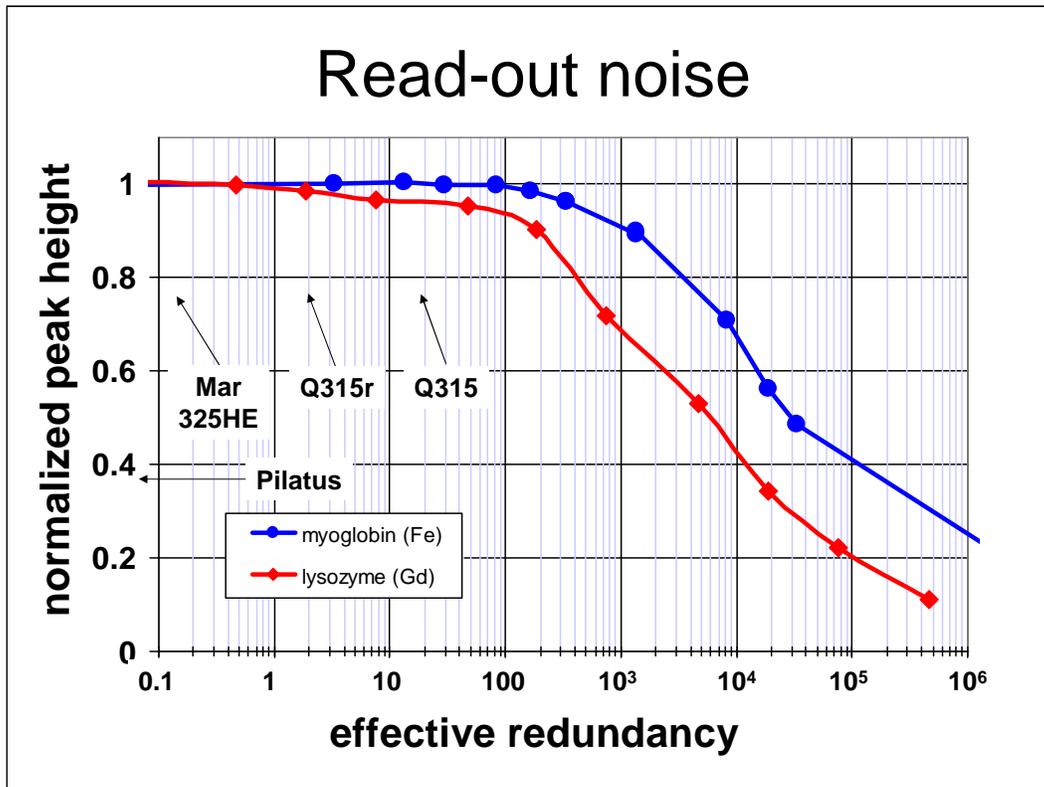
Increasing the signal strength is not always beneficial. In this simulation, using an occupancy for the Gd sites above unity simulates using a more powerful anomalous scatterer. In this case, the map correlation declines because this is a SAD experiment and the estimation of FH becomes less accurate as the FH vector becomes large. For a MAD experiment, the curve is flat at occupancy = 1 or higher.



The exponential decay of heavy atom sites is now a well-known hallmark of radiation damage, but the exact effect that decaying sites have on “solvability” has not been shown in the literature. In this simulation either one (green line) or both (blue line) of the Gd sites in the model were made to decay exponentially with dose and the “half dose” (dose at half reaction point) of this exponential was varied on the x axis. The red bar denotes the total dose of the complete data set. Note that one decaying site is only half as detrimental as two. This was unexpected since asymmetric site decay rates cause the phase of the heavy atom model to evolve with dose and this phase shift is not taken into account by MLPHARE. It would appear from this simulation that only total anomalous signal strength is important.



Detective quantum efficiency (DQE) is a highly prized feature of new detectors. The “DQE” varied in this simulation is the fraction of incident photons that contribute to the “counting noise” step. Fewer “detected” photons means that the signal/noise drops accordingly. Not that the simulation is highly insensitive to this DQE. The reason for this is that the DQE is highly related to exposure time (explored on an earlier slide). The “turnover point” on this graph corresponds to an effective exposure time of 0.01s, which is also the turnover point on the exposure time graph. It would appear from this simulation that a DQE close to unity is desirable, but a fractional drop in DQE from unity can be effectively compensated by increasing the exposure time by that same fraction.



Modern detector developers are endeavoring to push the contribution of intrinsic noise (read-out noise in the case of a CCD) to zero. One way to interpret read-out noise is as an “effective redundancy” since the detrimental effect of increasing redundancy with the same total exposure is to increase the total amount of read-out noise that contributes to the data. In this graph the effective redundancy was modeled by adding excess read-out noise to the images. Note that the “turnover points” occur when the read-out noise is made to be several orders of magnitude higher than is seen with modern detectors. As with most of these simulations, the “turnover point” occurs when the total noise added approaches the Bijvoet ratio. The Bijvoet ratio of the myoglobin data (1%) is lower than that of the lysozyme data (3.5%), which means that a higher level of read-out noise is required for that noise level to approach the scale of 1% of the mean value of  $F$ . The counterintuitive conclusion is that weaker anomalous signals are less sensitive to read-out noise than strong anomalous signals because weak anomalous signal measurements are dominated by photon-counting error, and not read-out noise.

<b>same total dose with high and low redundancy</b>			
dataset	1	2-11	12
exposure	1.0s	0.1s	1.0s
frames	100	100 x 10	100
$R_{\text{merge}}$	5.6%	<b>11.2%</b>	4.7%
$R_{\text{anom}}$	4.8%	4.7%	4.7%
I/sd	29.5	<b>43.4</b>	33.3
I/sd (2.0 Å)	23.3	<b>29.6</b>	25.8
redundancy	7.6	75.7	7.6
PADFPH	36.69	37.11	37.93
FOM	0.342	0.343	0.366
FOMDM	0.698	0.711	0.726
CC(1H87)	0.418	<b>0.492</b>	0.468

The expected insensitivity of MAD/SAD to read-out noise suggests that dividing a given total exposure over more images should not have a significant detrimental effect on data quality. This was tested experimentally. Here the same 100-image data set was collected 12 times, once with 1 s exposures, then ten times with 0.1 s exposures and again with 1 s exposures to test for radiation damage, which was not evident. Note that merging the 1000 frames with 0.1 s exposures resulted in a significantly higher  $R_{\text{merge}}$ , but the I/sd and the final correlation to the correct model are also higher than the non-redundant data sets. In this case, it would appear that  $R_{\text{merge}}$  is not a good indicator of data quality. The reason for the unexpected improvement in data quality by dividing the same number of photons over more images is under investigation, but it could be due to the improved stability of scaling algorithms as the observations/hkl becomes high.